

## APPENDIX A: Document Type Definition of Multimedia Integration Description Scheme

### MM\_integration\_ds.dtd

5

```
<!-- Multimedia Integration Description Scheme -->
```

```
<!ELEMENT MM_Stream ( MM_Object_Set, Object_Hierarchy*,
Entity_Relation_Graph* )>
```

10

```
<!ATTLIST MM_Stream
      id ID #IMPLIED>
```

```
<!-- This is how external DTDs are included in the current DTD -->
```

15

```
<!-- External Video DS DTD -->
```

```
<!ENTITY % Video_DS SYSTEM "video_ds.dtd">
%Video_DS;
```

```
<!-- External Audio DS DTD -->
```

20

```
<!ENTITY % Audio_DS SYSTEM "audio_ds.dtd">
%Audio_DS;
```

```
<!-- External Text DS DTD -->
```

25

```
<!ENTITY % Text_DS SYSTEM "text_ds.dtd">
%Text_DS;
```

```
<!-- External Synthetic DS DTD -->
```

30

```
<!ENTITY % Synthetic_DS SYSTEM "synthetic_ds.dtd">
%Synthetic_DS;
```

```
<!-- External Image DS DTD -->
```

```
<!ENTITY % Image_DS SYSTEM "image_ds.dtd">
%Image_DS;
```

35

```
<!ELEMENT MM_Object_Set ( MM_Object+ )>
```

```
<!ELEMENT MM_Object ( Media_Object_Set, Object_Hierarchy*,
Entity_Relation_Graph*,
```

40

```
      MM_Obj_Media_Features?, MM_Obj_Semantic_Features?,
      MM_Obj_Temporal_Features? )>
```

```
<!ATTLIST MM_Object
```

```
      Object_Type
```

```
      (LOCAL|GLOBAL)
```

```
      #REQUIRED
```

```

        id            ID            #IMPLIED
        Object_Ref    IDREF         #IMPLIED
        Object_Node_Ref IDREFS      #IMPLIED
        Entity_Node_Ref IDREFS      #IMPLIED>
5
<!ELEMENT Media_Object_Set ( Audio_Object | Image_Object | Video_Object |
Text_Object
                               Synthetic_Object )+ >

10
<!-- The object hierarchy and the entity relation graph are defined in the Image DS
(Proposal #480). We include them in this DTD for convenience -->
<!-- Object hierarchy element -->
<!-- The attribute type is the hierarchy binding type -->
<!ELEMENT Object_Hierarchy ( Object_Node )>
15
<!ATTLIST Object_Hierarchy
        id            ID            IMPLIED
        type          CDATA        #IMPLIED>

<!ELEMENT Object_Node ( Object_Node* )>
20
<!ATTLIST object_node
        id            ID            #IMPLIED
        Object_Ref    IDREF        #REQUIRED>

<!-- Entity relation graph element-->
25
<!-- Possible types of entity relations and entity relation graphs:
        - Spatial: topological, directional
        - Temporal: topological, directional
        - Semantic -->
<!ELEMENT Entity_Relation_Graph ( Entity_Relation+ )>
30
<!ATTLIST Entity_Relation_Graph
        id            ID            #IMPLIED
        type          CDATA        #IMPLIED>

<!ELEMENT Entity_Relation ( Relation, (Entity_Node | Entity_Node_Set |
35
Entity_Relation)* )>
<!ATTLIST Entity_Relation
        type          CDATA        #IMPLIED>

<!ELEMENT Entity_Node (#PCDATA)>
40
<!ATTLIST Entity_Node
        id            ID            #IMPLIED
        Object_Ref    IDREF        #REQUIRED>

<!ELEMENT Entity_Node_Set ( Entity_Node+ )>

```

```

<!ELEMENT Relation Temporal_Parallel| Temporal_Sequential |
                    Spatial_Alignment| Spatial_Arrangement |
                    Semantic_Relation| code)*>
5  <!ATTLIST   Relation
    With_Respect_To   IDREF   #IMPLIED >

<!ELEMENT   Temporal_Parallel   EMPTY >
<!ELEMENT   Temporal_Sequential EMPTY >
10 <!ATTLIST   Temporal_Sequential
    Pattern   (EXACT|DELAY|PRIOR)   "EXACT" >

<!ELEMENT   Spatial_Alignment EMPTY >
<!ATTLIST   Spatial_Alignment
15   Pattern (Left_Align| Right_Align |
    Top_Align| Bottom_Align)   "Left_Align"
    At_Time   CDATA   #IMPLIED >

<!ELEMENT   Spatial_Arrangement ( Spatial_Relevance| Spatial_Positioning )? >
20 <!ATTLIST   Spatial_Arrangement
    At_Time   CDATA   #IMPLIED >

<!ELEMENT   Spatial_Relevance EMPTY >
<!ATTLIST   Spatial_Relevance
25   Pattern (Top_Of| Bottom_Of| Left_Of| Right_Of|
    Upper_Left_Of| Upper_Right_Of| Lower_Left_Of|
    Lower_Right_Of |
    Adjacent_To| Neighboring_To| Near_By |
    Within| Contained_In )   "Top_Of" >

<!ELEMENT   Spatial_Positioning EMPTY >
30 <!ATTLIST   Spatial_Positioning
    Horizontal_Shift   CDATA #IMPLIED
    Vertical_Shift     CDATA #IMPLIED >

35 <!ELEMENT   Semantic_Relation (Keywords*| code*)? >

<!ELEMENT MM_Obj_Media_Features ( Data_Location?, Scalable_Representation?,
                                Modality_Transcoding? )>
40 <!ELEMENT MM_Obj_Semantic_Features ( Text_Annotation?, Keywords? )>

<!ELEMENT MM_Obj_Temporal_Features ( Duration? )>

```

5 <!ELEMENT Keywords ( Word\*, Code\* ) >  
 <!ATTLIST Keywords  
     No-Words CDATA #REQUIRED  
     Language CDATA "English"  
     Extraction\_Manner (AUTOMATIC|MANUAL) "MANUAL" >

10 <!ELEMENT Duration (  
     Image\_Duration?,  
     Audio\_Duration?,  
     Video\_Duration?,  
     Text\_Duration?,  
     Synthetic\_Duration? ) >  
 <!ATTLIST Duration  
     Synchronized\_Overall\_Duration CDATA #IMPLIED >

15 <!ELEMENT Image\_Duration ( Time ) >  
 <!ELEMENT Audio\_Duration ( Time ) >  
 <!ELEMENT Video\_Duration ( Time ) >  
 <!ELEMENT Synthetic\_Duration ( Time ) >  
 <!ELEMENT Text\_Duration ( Time? | Alignment? ) >  
 20 <!ELEMENT Alignment EMPTY >  
 <!ATTLIST Alignment  
     with (IMAGE|AUDIO|VIDEO|SYNTHETIC) "AUDIO" >

25 <!ELEMENT Data\_Location(  
     Image\_Location?,  
     Audio\_Location?,  
     Video\_Location?,  
     Text\_Location?,  
     Synthetic\_Location?) >  
 30 <!ELEMENT Image\_Location (location) >  
 <!ELEMENT Audio\_Location (location) >  
 <!ELEMENT Video\_Location (location) >  
 <!ELEMENT Text\_Location (location) >  
 <!ELEMENT Synthetic\_Location (location) >  
 35

<!ELEMENT Scalable\_Representation (  
     Static\_Sampled?,  
     Dynamic\_Condensed?) >

40 <!ELEMENT Static\_Sampled (  
     Image\_Condensed?,  
     Video\_Static\_Pictures?,  
     Audio\_Clips?,

```

Synthetic_Pictures?) >
<!ELEMENT Image_Condensed ( Location*, Image_Scl*, Image_Subsampling* ) >
<!ELEMENT Image_Subsampling ( Image_Subsampling_Para, code)* >
<!ELEMENT Image_Subsampling_Para EMPTY >
5 <!ATTLIST Image_Subsampling_Para
    Scheme CDATA #REQUIRED
    Spatial_Rate CDATA #REQUIRED
    Frame_Size CDATA #IMPLIED >
<!ELEMENT Video_Static_Pictures (Key_Frame* ) >
10 <!ATTLIST No_KFs CDATA #REQUIRED >
<!ELEMENT Audio_Clips (Audio_Object* | Audio_Hierarchy*) >
<!ATTLIST No_Clips CDATA #REQUIRED >
<!ELEMENT Synthetic_Pictures (Key_Frame)* >
<!ATTLIST No-KFs CDATA #REQUIRED >
15
<!ELEMENT Dynamic_Condensed (
    Visual_Condensed?,
    Audio_Condensed?,
    Text_Condensed?,
20 Synthetic_Condensed?) >

<!ELEMENT Visual_Condensed ( Location*, Video_Scl*, Video_Subsampling* ) >
<!ELEMENT Video_Subsampling ( Video_Subsampling_Para, code)* >
<!ELEMENT Video_Subsampling_Para EMPTY >
25 <!ATTLIST Video_Subsampling_Para
    Scheme CDATA #REQUIRED
    Temporal_Rate CDATA #IMPLIED
    Spatial_Rate CDATA #IMPLIED
    Frame_Size CDATA #IMPLIED >
30 <!ELEMENT Audio_Condensed ( Location*,
    Audio_Compressed*,
    Audio_Subsampling*,
    Audio_Timescaled* ) >

<!ELEMENT Audio_Compressed ( Audio_Compress_Para, code)* >
35 <!ELEMENT Audio_Compress_Para EMPTY >
<!ATTLIST Audio_Compress_Para
    Scheme CDATA #REQUIRED
    Bitrate CDATA #IMPLIED >

<!ELEMENT Audio_Subsampling ( Audio_Subsampling_Para, code)* >
40 <!ELEMENT Audio_Subsampling_Para EMPTY >
<!ATTLIST Audio_Subsampling_Para
    Scheme CDATA #REQUIRED
    Temporal_Rate CDATA #IMPLIED >
<!ELEMENT Audio_Timescaled (Audio_Timescale_Para, code)* >

```

<!ELEMENT Audio\_Timescale\_Para EMPTY >  
 <!ATTLIST Audio\_Timescale\_Para  
           Scale\_Rate       CDATA #REQUIRED >

5   <!ELEMENT Text\_Condensed (Text\_Abstract\*) >  
     <!ELEMENT Text\_Abstract (Location\*, (Text\_Abstract\_Para, code)\*)? >  
     <!ATTLIST Text\_Abstract  
                 Length\_In\_Words       CDATA                   #IMPLIED  
                 Duration\_In\_Seconds   CDATA                   #IMPLIED  
 10               Language            CDATA                   "English"  
                 Generation\_Mode   (AUTOMATIC|MANUAL) "MANUAL" >  
     <!ELEMENT Text\_Abstract\_Para EMPTY >  
     <!ATTLIST Text\_Abstract\_Para  
                 Length\_In\_Words       CDATA                   #IMPLIED  
 15               Language            CDATA                   "English" >

    <!ELEMENT Synthetic\_Condensed (Synthetic\_Location\*, (Synthetic\_Condense\_Para,  
     code)\*) >  
     <!ELEMENT Synthetic\_Condense\_Para EMPTY >  
 20    <!ATTLIST Synthetic\_Condense\_Para  
                 Spatial\_Rate        CDATA                   #REQUIRED  
                 Temporal\_Rate       CDATA                   #REQUIRED  
                 Frame\_Size          CDATA                   #REQUIRED  
                 Bitrate             CDATA                   #IMPLIED >

25   <!-- Multimedia Integration DS End -->